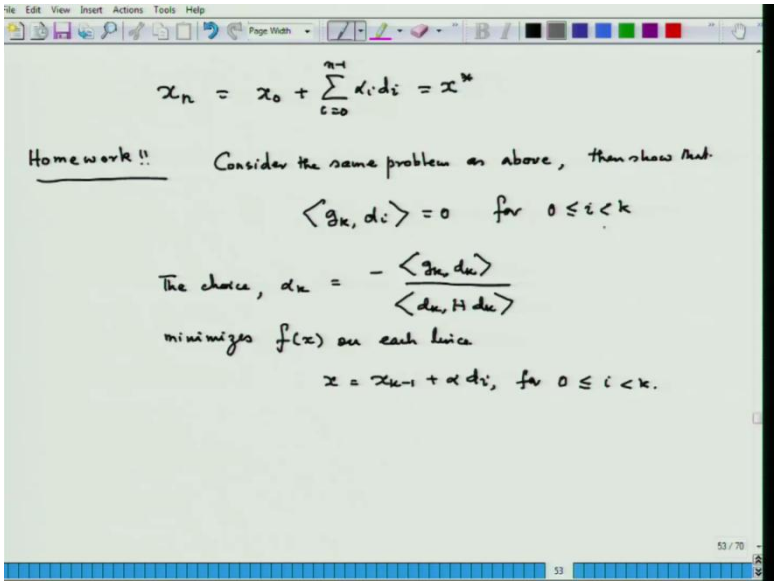
Foundation of Optimization
Prof. Dr. Joydeep Dutta
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

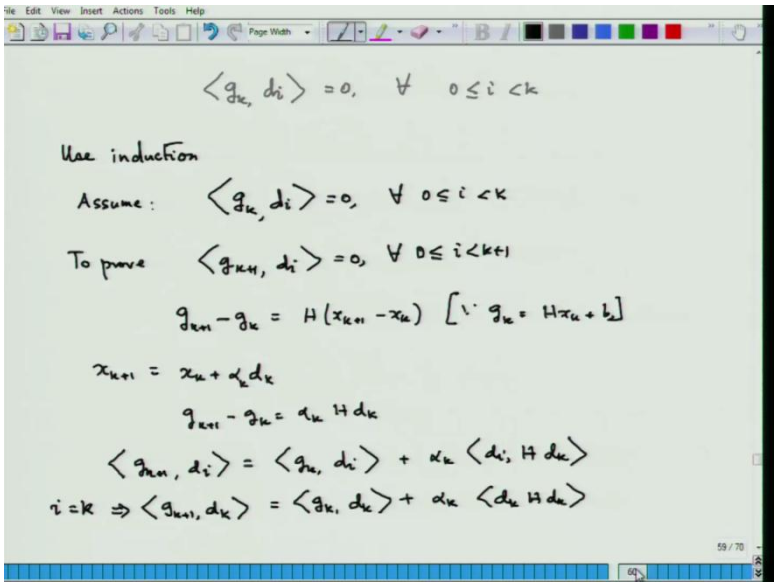
Lecture – 12

(Refer Slide Time: 00:25)



Now, I given you this homework while discussing the conjugate directions method in some of the, in one of the previous lectures. Now I will solve this part of this homework this part, this part, but I will keep this for you to try out. So, today I will begin by trying to solve. So, what I have to prove is the conjugate directions and the gradients are perpendicular to each other.

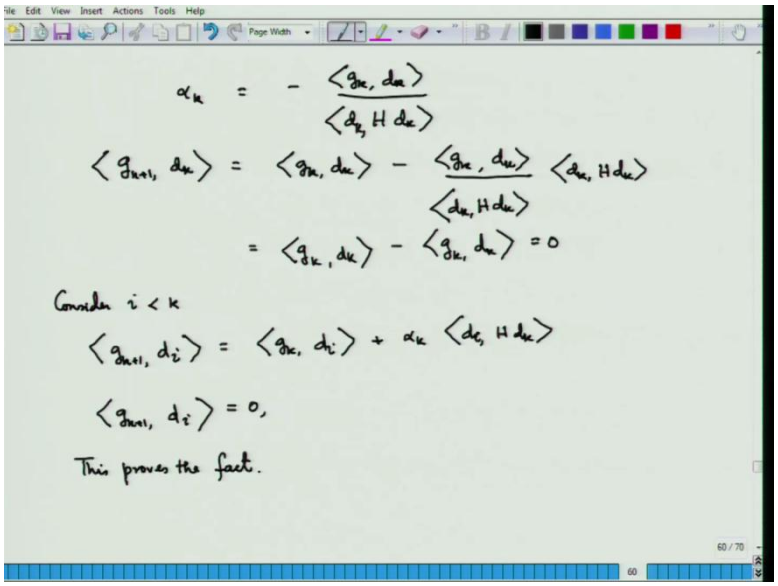
(Refer Slide Time: 00:56)



But, so what I have to really prove is a following is g k, d i at any k is equal to 0 for all i for, for all i which is bigger than equal to 0 or less than or equal to k strictly. What we do is that we use induction. So, by induction what we show is a following that we show that let us assume g k, d i is equal to 0 for all and to prove g k plus 1 d i is equal to 0 for, so that is exactly what we have to do. So, you we have assume this now let us start working it out; let us observe this fact. Now this one is very simple since g k is nothing but, H x k plus b. So, g k plus 1 would be nothing but H x k plus 1 plus b, so b would get cancelled out and this is what you will have.

Now of course, you know that x k plus 1 is x k plus alpha that is how you update using the conjugate gradient direction that is how you update alpha k d k. So, then g k plus 1 minus g k, now I can write this as alpha k d k, so it will become alpha k H d k. Now I multiply both sides by (( )) take the inner product both sides by d i right for i. So, then I will get g k plus 1, d i is equal to g k, d i plus alpha k, d i H d k. Now I have to prove this fact. So, basically I have to prove that for all 0 for all i equal from 0 to i equal to k this thing holds. So, first put i is equal to k; imply that plus alpha k d k H d k. You know what is alpha k, alpha k is already you have solved out alpha k. Now what we have to do is now here I have to put the value of alpha k.

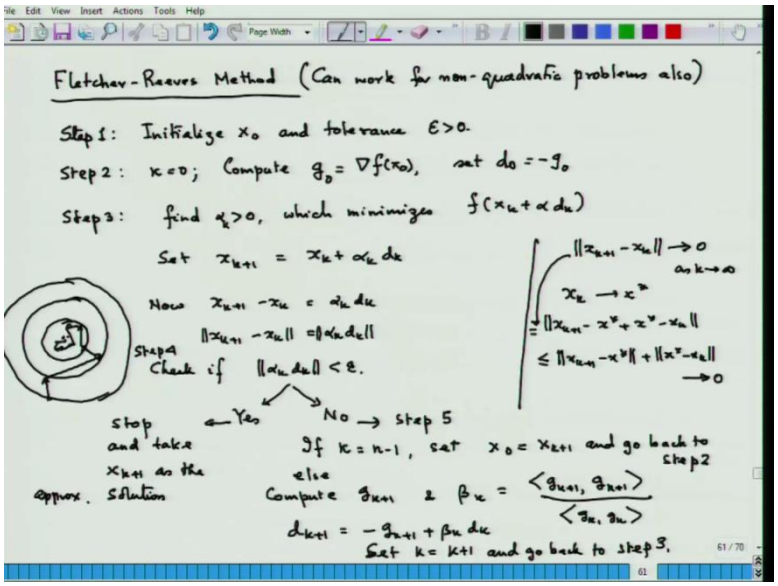
(Refer Slide Time: 04:15)



So, alpha k is already known to me and that is minus g k, d k, d k, H d k. So, if you do that, so then g k plus 1, d k is nothing but g k, d k minus orientation, so that can following it - g k, d k. This and now here also we have d k H d k which cancels up now H is positive definite matrix, so it will this d k is non-zero, so these are all positive. So, this will cancel up and so we will be left with g k, d k. Now once you have done, this is 0. So, you have proved for k, now you have to prove for anything other than k. Now take consider i is strictly less than k then for that g k plus 1, d i is g k, d i plus alpha k I think alpha k you will come here d I, H d k. So, this i is strictly less than k now, so our d i is not i is not equal to k. So, then by the fact that these are conjugate directions this would become 0 and the fact that we have already assumed when we started the proof then this would also become 0.

So ultimately, so this proves the fact. Now what is important to know that all these things that we have done, all the conjugate direction or conjugate gradient method have been really applicable for convex problem that to with H a convex quadratic problem with H positive definite is really strongly convex or strictly convex quadratic problems. What about handling it for non-quadratic problems or anything it is a for any sort of convex problems for example, can we do something with that.

(Refer Slide Time: 07:23)



So, for that Fletcher reeves introduced a method called the Fletcher Reeves method. We are just going to outline this method here. Now let us write down the Fletcher Reeves method step by step. This can work for non-quadratic problems also. Actually the shift from linear to non-linear is quite a difficult shift methods you know they are not once you have non convex problems you have very less things available to you. You just make a trial with whatever algorithms you have in your hand and see what you have that that is the only thing that you can possibly do. And recent research there are efforts to take those points and try to give some quantitative justification about their behavior, but that is absolutely at the frontier of research.

So, we would not get into that issue at all, but I just want to recall and remind you that non convex optimization by the way is very very hard, of course, there could be convex optimization problems which are also not so easy to handle, but there are algorithms which will support them. But for non convex problems, you just do not know in large cases what to do. We will learn about sequential quadratic programming method methodology as we go along, but you see that that has its own drawbacks as we come to it.

Now in Fletcher reeves method, so step one is to initialize your starting point x 0 and tolerance epsilon. Step two is to set k is equal to 0; step two, you should compute g 0 that is the gradient at x 0 and set the first conjugate direction that is set. Step three - it was the most idealistic stuff find alpha greater than zero, which minimizes of course this

minimization is done in a approximate way. So, here you have one dimensional minimization, we have not spoken much about one dimensional minimization in this course, possibly at the very end of this study about unconstraint optimization, we will take up a one day to explain some very important one dimensional minimization techniques. Because when you want to solve this problem where x k and d k are fixed then this is nothing but a function in alpha. So, then basically you are talking about one dimension minimization minimization of a one of a real function in real variables, so f from r to r.

Now once we have done that, so that is alpha k. So, basically find alpha k I should say which minimizes this set your next iterate to x k plus 1 and x k plus 1 is x k plus alpha k d k is that I cannot I really do not know the solution. But when the distance between x k plus 1 and x k - these distance that distance comes becomes very very small that is we are basically coming near the solution. Like if you had thought about steepest descent, so there was this is your this is your level curves, suppose say what would happen level curves means these are the function values say f x y equal to c, because we are in two dimensional set up you can see it very well. So, you are here, so you take one direction and you moved here then you took another direction and you moved here then you are moving there then you are moving there and then you are moving there and then here then here then here then here. So, as go near the solution, these distance also decreases.

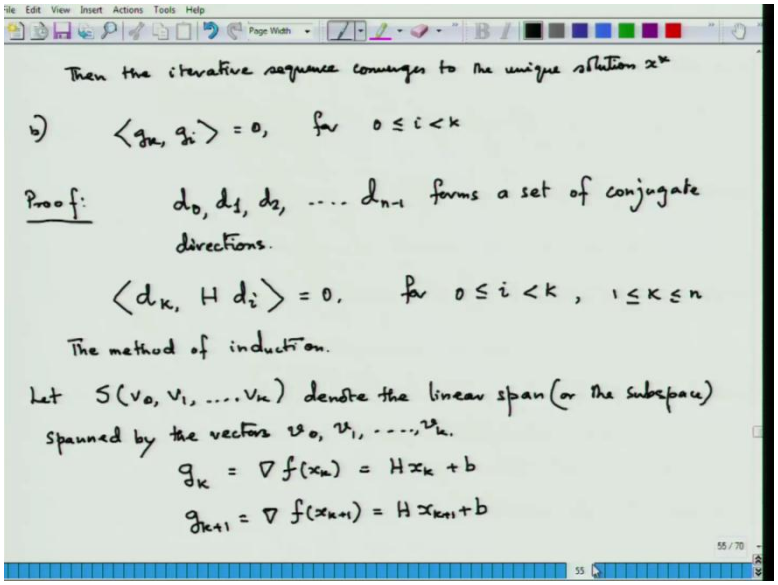
Let me also give an explanation that so if you thing is actually taking you to the solution then this is what should happen that is if x k is going to the solution x star then this is what should happen. Because then what I can do is that I can write x k plus 1 minus x star this is equal to plus x star minus x k, and this is nothing but less than x k plus 1 minus x star plus x star minus x k. And this all also this whole thing now goes to 0. So, if I can show that the distance between these two consecutives one's becomes very very small, basically I am trying to show that it is a essentially in some sense. So, if I can show that then I know that it will converge that it will actually go to towards the solution. So, for sufficient for k sufficiently large, we can show that this is true this distance is very very small then I am almost near the solution then I can stop there.

So, what I do now x k plus 1 minus x k is equal to alpha k d k, and norm x k plus 1 minus x k is equal to alpha k d k. So, basically you really do not have to bother about x k plus 1 at very first, you just take alpha k d k check if norm alpha k d k is strictly less than epsilon - step 4. So, there are two answers to it yes and no. So, if it is yes, stop and take x k plus 1

as the solution approximately, as the Approximate solution I should write. If it is no, then do the following there comes step five. If k is equal to n minus 1, set x 0 is equal to x k plus 1 and go back to step 2. If k is equal to n minus 1, so we have not reached our solution in n minus 1 steps then we have to again restart the procedure. If not else compute g k plus 1 and beta k is equal to g k plus 1 g k plus 1 g k g k that is nice one and then you compute the next direction d k plus 1 because from there the x k plus 1 now you have to go to x k plus 2. So, you are computing the d k plus 1 is equal to as minus g k plus 1 in the same way we have done for the power this hastiness conjugate gradient method plus b k to d k.

And you see now once you have that what you can do is set k is equal to k plus 1 and go back to step three or rather assign k equal to k plus 1. So, this is an approach see why this k equal to n minus 1, because in the quadratic case when you have positive definite hessian in n minus 1 steps you are basically getting the solution. So, in n iterations basically n iterations you are getting the solution, and here I am in n minus 1 eth step and I am yet to get the solution. So, I have to reset x naught as x one and taking up the starting point and start the whole procedure again. Because at k is equal to n minus 1, I have not got the solution this is some sort of slight restriction that is there, but k equal to n, I am suppose to get the solution and I started from x 0. So, I have x 0 then in n steps, so x n minus 1 would be my just a moment…

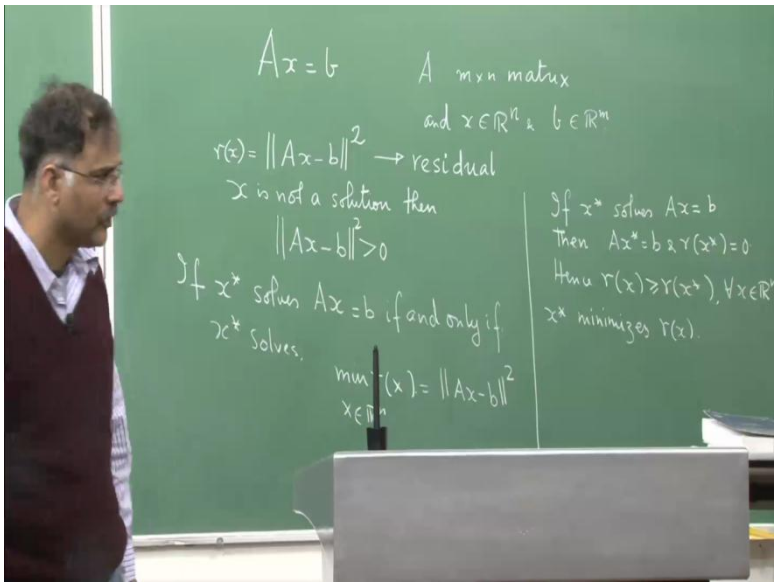
(Refer Slide Time: 18:25)



So, as we have already seen in our previous studies that in n steps in the case of when H is positive definite, in n steps we are coming to the solution. So, this we can come to the solution in just n steps. We start with x 0, and x n would be x star. So, here when k is you can say, why k should not be n minus 1, but I have started with k equal to 0 and I have come to k equal to n minus 1 which I have, so I have taken n steps right 0 th step first step second step third step n eth step. n minus 1 is the n eth step in this case. So, I have reached the n eth step. So, even if I have reached the n eth step, so x n should be my solution. So, here I have reached the n eth step, but I have not got the solution that that is the if k is that that is the whole idea if k is equal to n minus 1. So, I have already taken n eth k is from 0 to n minus 1, I have taken n step, this n eth step and 0 th iteration the first iteration 0 th iteration first iteration second iteration k n minus 1.

So, basically I have taken n steps I have got my come to my n minus 1 th iteration in next iteration, I am suppose to get the solution. But at k equal to n minus 1, if I do not get the solution then I really have to restart the procedure that because we are handling non convex problems, we are not sure whether at nth step we are going to get a solution. We have non-quadratic problem with we do not know anything about the of functions nature. So, then a little bit of extra caution is kept here by putting k is equal to n minus 1 and then again restarting the whole procedure from x not equal to x k plus 1. Now we are going to end our discussion of conjugate directions and conjugate gradient method which is very important class and then we will go into in a very slightly interesting problem called the least square problem. Now what does this least square problem means? So, least square problems is how optimization can help you in solving, how optimization can help you in solving equations.

(Refer Slide Time: 21:08)

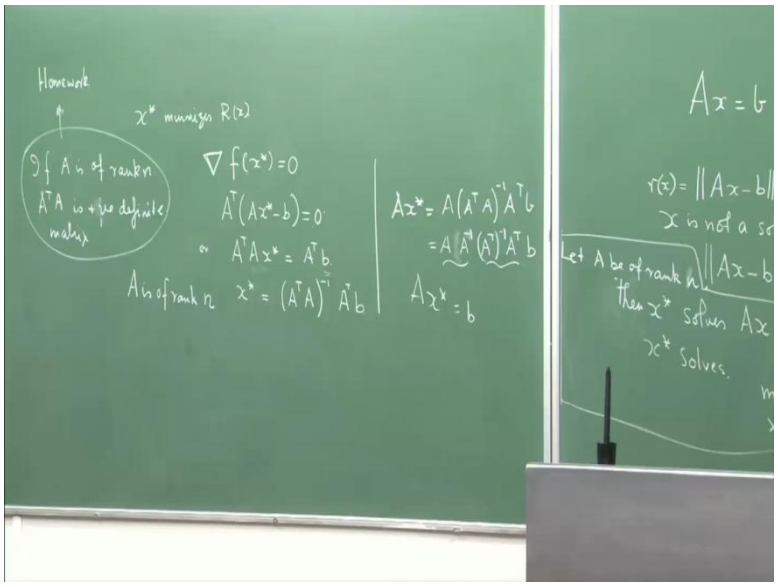


And let us just see what can we do about it one to I have a m cross n matrix, and I want to solve this equation. So, A is m cross n matrix, and x is a n vector, and b is in R m . So, now, this need not have unique solution, it depend on the relation between m and n. So, in general, it can have many solutions, and it is not see I do not know that is a there is a because m is not equal to n, and I have no idea about the inevitability in this case. And, so I cannot really figure out what is the solution so easily, it is not so easily to figure out one solution even. If b is 0 then x can be 0 then one of the solution; if b is non 0 I do not know. How do I try to attempt to solve this sort of system of equation, because these things come of very much in applications. See what I can prefer to do instead of trying to solve it, I construct this function is called the residual function. So, if I take n x, I take any x in R n and put here in and multiply it with A, then if A x is not equal to b if it is not the solution then A x minus b this vector is a non zero vector, and then the norm of that would be non zero.

So, if x is not a solution, then so this quantity is called a residual, this quantity sometimes for the more terminology loving b person is called a residual. So, x is not a solution then this will happen. So, if x is a solution A x minus b is would be equal to 0, and then that x would actually be the minimum of this problem, because this is always greater than equal to 0. For any x for which A x minus b is a x is A x is equal to b that x must be a solution of the minimization of this problem over R n. So, if x star solves A x equal to b, if and only if x star solves the problem minimize r x, x element in R n where r x is in this case now how

do you… So, if I want to solve this problem, I can actually solve the this minimization problem, but how do I prove that. If x star solves A x equal to b then naturally A x star is equal to b and r x star is equal to 0, hence r x is greater because r x is greater than equal to 0 for all x, so r x star, so x star minimizes r x. Now if x star minimizes r x, how do I show that you have this as a solution, if x star minimizes r x then that x star would solve A x star equal to b.

(Refer Slide Time: 26:24)



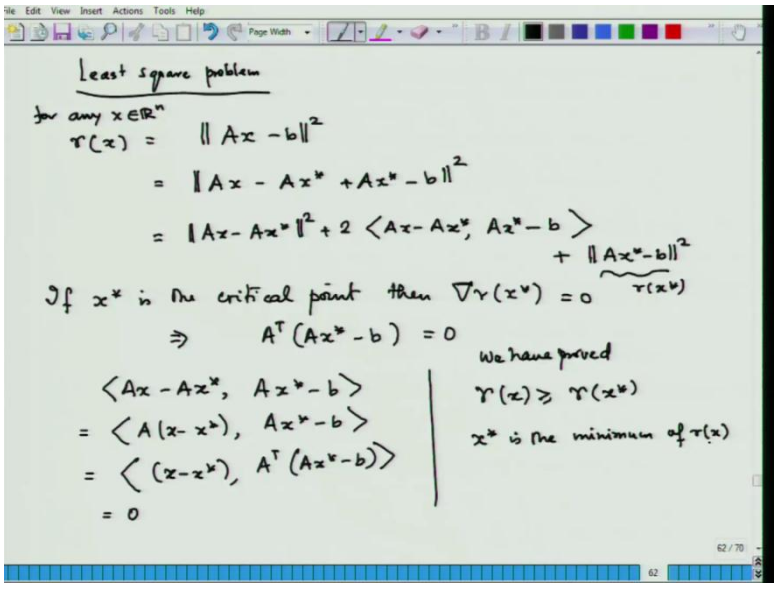
So, if x star minimizes suppose x star minimizes r x then what would happen, I can write that the gradient of this problem must be 0. So, I have already know that x star is minimizing r x. So gradient of f x star, so that would give me A transpose A x star minus b is equal to 0 or A transpose A x star is equal to A transpose b that is what it gives me. What I have found here is actually a critical point, so if this happens then this is what will happen, then can I say here that A x star is equal to b from this can I say that A x star is equal to b which looks quite clear. Now suppose A is of full rank, A is of full column rank then A transpose A is actually invertible then you can write x star if a is of rank n that is a is of full column rank if a is of rank n then x star is equal to A transpose b.

Now if I can get this, so if x star solves this problem then grad f x star must be equal to 0 and x star must have this form. Then if A is of rank n then x star must have this form then what I have if a x star then A of x star is equal to A of A transpose A inverse A transpose b. See because A is of rank n, A transpose A becomes a positive definite matrix that is

very important and that is why it is invertible. So, if A is of rank n is a positive definite is a positive definite matrix. So, we prove this in the homework, so this is your homework. Now what is A x star. So, let me see what happens. So, now, this can be written as A A inverse A transpose inverse where A b inverse is b inverse A inverse. So, this is A inverse A transpose inverse A transpose inverse b, this is identity and this is identity. So, you have A x star is equal to b. So, I will now modify what I have written.

So, this is how you discover things in mathematics by trying it out. So, I have not written it purposefully, but what I have now done what I have written, I have written that if A is of rank n, let so my result is let a be of rank n of full column rank. Then x star solves A x equal to b if and only if x star solves this problem. Then so you see then now we have the full result now this is my result. So, basically, now if I want to solve this problem, what I will first do is I will first find a point like this, and then if A is of rank n, and if I am now trying to minimize this problem. If I find a critical point of this problem, any critical the critical point of x star if A is of rank n critical point of r x I have written should be r x critical point of r x is of this form. And we want to show that this critical point is actually a minimum this critical point is actually a minimum that is very very important to show.

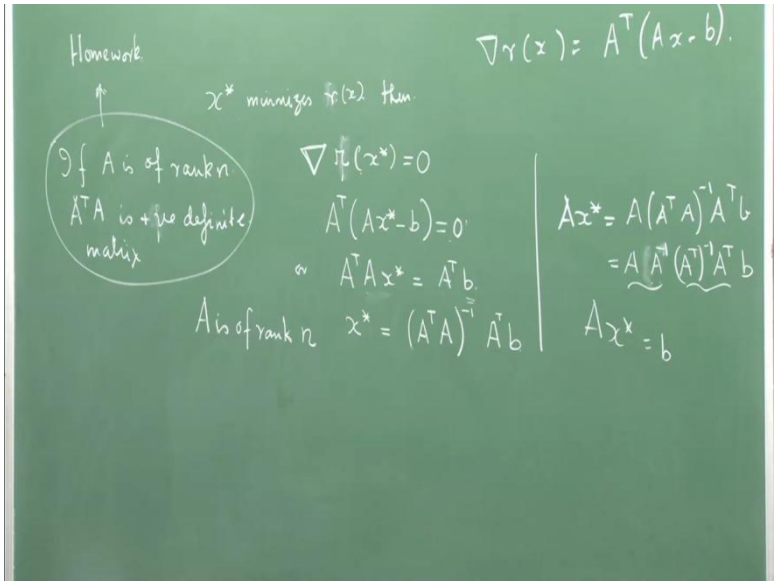
(Refer Slide Time: 32:50)



So, r of x is norm A x minus b whole square. So, take any x in R n, now once you know this, you can write this as x star minus b. Now this term would be 0, because of this fact because we know that if x star is a critical point, so the critical point of this least square

problem if I minimize this if I find this critical point when a rank of A is n then that critical point is actually a solution of this problem. And we have shown that that critical point is actually solving the original problem A x star equal to b. So, it is very nice that I can actually in an exact way I can find I mean exact analytic expression for this and that is a beautiful part which you do not always have in optimization. So, if x star is a critical point in then what you have, so the derivative I have again made a small mistake in my writing, but please forgive me for this, because this is something you can figure out because here instead of A it should be r, because we are trying to find the critical point of r.

(Refer Slide Time: 35:38)



So, then grad r x star is equal to 0 and again I will leave it as a homework for you to figure out that the derivative of r x that is it is your homework to figure out that the gradient of r x is nothing but A transpose A x minus b. So, this implies A transpose A x star minus b is equal to 0. Now if you look at the expression 2 or A x minus A x star. So, let me just forgot out this two. So A x star minus b and this is nothing but A into x minus x star while matrix property and A x star minus b and that is equal to x minus x star A transpose A x star minus b and that is you already know it is 0, so this is 0 . So, now, this is equal to 0 and this is greater than equal to 0, so what we have proved that we have proved that r x is greater than equal to r x star is this part.

So, this is your r x star. So, hence x star is the minimum of r x. So, any critical point, there is one critical point, the critical point of this function is actually minimum of this and

hence and is also a solution of the equation A x star A x equal to b provided A is of rank n. So, this is a very very important requirement of rank n. So, it is some little bit of extra things that we have if you put in some little assumption things flow in a much more interesting way. So tomorrow we will start in the next class by talking to you about the Gauss Newton method and talking to you about the least square problem in a more general way. This is just an example to show you the importance of the least square problem, how optimization can be used even to solve a equation this linear equation. So, and then we will discuss a bit about Gauss Newton method and after that we will start talking about the Quasi-Newton method and then return to theory for a while for and then get into return to the theory for a while for the unconstrained case and get into the study of the celebrated Kuhn Tucker conditions.

Thank you.